



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/777,072	02/13/2004	Soo Jun Park	123056-05004506	4141

22429 7590 02/04/2008  
LOWE HAUPTMAN HAM & BERNER, LLP  
1700 DIAGONAL ROAD  
SUITE 300  
ALEXANDRIA, VA 22314

EXAMINER

SIDLER, DOROTHY S

ART UNIT	PAPER NUMBER
----------	--------------

2626

MAIL DATE	DELIVERY MODE
-----------	---------------

02/04/2008

PAPER

**Please find below and/or attached an Office communication concerning this application or proceeding.**

The time period for reply, if any, is set in the attached communication.

<b>Office Action Summary</b>	Application No. 10/777,072	Applicant(s) PARK ET AL.	
	Examiner Dorothy Sarah Siedler	Art Unit 2626	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

### Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See, 37 CFR 1.704(b).

### Status

- 1) ☒ Responsive to communication(s) filed on 20 November 2007.
- 2a) ☒ This action is **FINAL**.                      2b) ☐ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

### Disposition of Claims

- 4) ☒ Claim(s) 1-15 is/are pending in the application.
- 4a) Of the above claim(s) \_\_\_\_\_ is/are withdrawn from consideration.
- 5) ☐ Claim(s) \_\_\_\_\_ is/are allowed.
- 6) ☒ Claim(s) 1-5 and 9-15 is/are rejected.
- 7) ☐ Claim(s) 6-8 is/are objected to.
- 8) ☐ Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

### Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on \_\_\_\_\_ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.  
     Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).  
     Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

### Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).  
     a) ☐ All    b) ☐ Some    \* c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
  2. ☐ Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.
  3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

\* See the attached detailed Office action for a list of the certified copies not received.

### Attachment(s)

- |  |   |
|--|---|
| 1) <input type="checkbox"/> Notice of References Cited (PTO-892)   | 4) <input type="checkbox"/> Interview Summary (PTO-413)<br>Paper No(s)/Mail Date. _____ |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948)                       | 5) <input type="checkbox"/> Notice of Informal Patent Application                       |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO/SB/08)<br>Paper No(s)/Mail Date _____ | 6) <input type="checkbox"/> Other: _____  |

## DETAILED ACTION

### *Response to Arguments*

Applicant's arguments filed November 20, 2007 have been fully considered but they are not persuasive.

Applicant argues that, "Johnson fails to disclose a rule collection unit configured to generate rules used to recognize the named entity" (Remarks page 7) as recited in claim 1, however the examiner respectfully disagrees. The applicant additionally states that, "Johnson, on the other hand, appears to disclose a method of constructing a special dictionary using rules designed to minimize the occurrence of multiple-semantic types by creating rules to diminish 'multiple-semantic type' (see page 211 and 213)" (Remarks page 7). The examiner agrees with the previous description, but adds that this special dictionary, or semantic lexicon as used in *Johnson*, is created from existing UMLS sources and subsequently used to aid in natural language processing (Abstract, page 205). This semantic lexicon is used to analyze and extract information from medical documents, including populating a database using information extracted from text reports (recognizing a named entity) (page 207, column 1 last paragraph and column 2 last paragraph).

Applicant also states that, "unlike Applicant's apparatus and method that initially builds a rule database based upon the databases derived from the UMLS, Johnson's 'semantic preference rules' are generated using both the semantic lexicon and the corpus of discharge summaries" (Remarks page 7); however the examiner notes that

the semantic lexicon of **Johnson** is derived from the UMLS, as stated in the Abstract (last paragraph, first sentence).

Applicant also argues that, "the Examiner respectively equates Johnson's terms 'lexeme' and 'semantic type' with Applicant's 'single name' and 'concept name' to conclude that the claimed databases are disclosed by Johnson. This is at least wrong in that the 'semantic type' as disclosed by Johnson corresponds to the classification of the meaning of some terms (see page 211, table 5), which is different from Applicant's disclosed 'concept name'" (Remarks page 8). However the examiner respectfully disagrees and notes that Applicant has simply provided a mere allegation of patentability, in that neither specific examples nor sufficient explanation was provided differentiating the terms used in the claim over the terms used in the prior art.

In response to applicant's argument that the references fail to show certain features of applicant's invention, it is noted that the features upon which applicant relies (i.e., the claimed keyterm is a portion of the claim concept name (Remarks page 8)) are not recited in the rejected claim(s). Although the claims are interpreted in light of the specification, limitations from the specification are not read into the claims. See *In re Van Geuns*, 988 F.2d 1181, 26 USPQ2d 1057 (Fed. Cir. 1993).

***Claim Rejections - 35 USC § 102***

The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless -

(b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

Claims 1-5, 9 -12, 14 and 15 are rejected under 35 U.S.C. 102(b) as being anticipated by ***Johnson*** ("A Semantic Lexicon for Medical Language Processing" JAMIA 1999).

1. As per claim 1; ***Johnson*** discloses an apparatus for recognizing a biological named entity from biological literature based on united medical language system (UMLS), comprising:

A resource construction unit for receiving metathesaurus from the UMLS and constructing a concept name database, a single name database and a category keyterm database; which are language resources to be used to recognize a named entity (page 211, Lexical Matching, *each lexeme in the specialists lexicon is matched to terms in the metathesaurus. Once a matched has been made, the lexeme (single name), semantic type (concept name) and all derivational and inflectional variants (category keyterm) are obtained and added to the semantic lexicon (database);*

A rule collection unit configured to receive each concept name stored in the concept name database, extracts a features from each of the concept names by using data stored in the single name database and the category keyterm database, and construct a rule database using the extracted features (page 211-213, *if one member of a pair of semantic types (concept name) is preferred for lexical items, including variants, (single names and keyterms) assigned to that pair, then a preference rule is determined. The rule is then assigned to each lexeme and variant in the semantic lexicon*);

A literature input configured to receive a biological literature (page 210 and 211, *Methods, the semantic lexicon is designed for analysis of discharge summaries (biological literature), therefore it is inherent that the system has a literature input*); and

A named entity recognition unit configured to receive the biological literature from the literature input, and extract candidate named entities from the biological literature and recognize named entities based upon the rules generated by the rule collection unit (page 211, *Corpus Matching, Contiguous word sequences were extracted from a corpus of discharge summaries and matched against the semantic lexicon*).

2. As per claim 2, **Johnson** discloses the apparatus of claim 1, wherein the resource construction unit extracts concept names from the metathesaurus of the UMLS, which is divided according to the semantic categories, to construct the concept names database, processes the concept name stored in the concept name database to

extract single names and category keyterms, and constructs the single name database and the category keyterm database by using the extracted single names and category keyterms (page 211, Lexical Matching, *each lexeme in the specialists lexicon is matched to terms in the metathesaurus. Once a matched has been made, the lexeme (single name), semantic type (concept name) and all derivational and inflectional variants (category keyterm) are obtained and added to the semantic lexicon (database);*

3. As per claim 3, **Johnson** discloses the apparatus of claim 1, wherein the rule collection unit extracts the feature of a token constituting each of the concept names stored the concept name database, creates the rules by combining the extracted features, weights the rules, filters the weighted rules with a threshold, and stores the filtered rules in the rule database (page 211-213, *lexemes (single names) and their inflectional variants (keyterms) are determined for a pair of semantic types (concept names) by comparing the semantic lexicon to the metathesaurus. Then discharge summaries are examined to determine which semantic type is used more frequently (weights). The semantic type that is used more frequently is preferred (filtered), and thus assigned to the lexeme and inflectional variants in the semantic lexicon).*

4. As per claim 4, **Johnson** discloses the apparatus of claim 1, wherein the named entity recognition unit extracts the candidate named entities from the literature provided through a literature input unit, extracts the feature of each of the tokens constituting the

candidate named entity, creates a rule used to determine the candidate named entity by combining the extracted feature, compares the created rule with the rule stored in the rule database to extract an existing rule suitable for the candidate named entity, applies a weight value of each of the extracted rules and a heuristic used to determine a category of the named entity, determines a final semantic category for the candidate named entity, and recognizing the named entity (page 211, Corpus Matching, and page 216, Results, *Contiguous word sequences were extracted from a corpus of discharge summaries and matched against the semantic lexicon. The semantic lexicon contains lexemes and semantic types combined to make a rule; therefore a rule must have been extracted from the literature in order to compare it to the semantic lexicon. In addition, preference rules (weight) are applied to the corpus to determine the semantic type.*

5. As per claim 5, **Johnson** discloses a method for recognizing a biological named entity from biological literature based on UMLS, the method comprising the steps of:

(a) receiving metathesaurus from the UMLS, (b) extracting concept names, single names and category keyterms and (c) constructing a concept name database, a single name database and a category keyterm database, (d) constructing a database of rules based upon information stored within the concept name database, the single name database, and the category keyterm database (page 211, Lexical Matching, *each lexeme in the specialists lexicon is matched to terms in the metathesaurus. Once a matched has been made, the lexeme (single name), semantic type (concept name) and*

*all derivational and inflectional variants (category keyterm) are obtained and added to the semantic lexicon (database) and page 211-213, if one member of a pair of semantic types (concept name) is preferred for lexical items, including variants, (single names and keyterms) assigned to that pair, then a preference rule is determined. The rule is then assigned to each lexeme and variant in the semantic lexicon);*

(e) inputting a literature (page 210 and 211, Methods, *the semantic lexicon is designed for analysis of discharge summaries (biological literature), therefore it is inherent that the system has a literature input*);

(f) extracting candidate named entities from the literature, and (g) recognizing named entities from the candidate named entities based upon the rules applied against the single name and category keyterm databases (page 211, Corpus Matching, *Contiguous word sequences were extracted from a corpus of discharge summaries and matched against rules in the semantic lexicon*).

6. As per claim 9, **Johnson** discloses the method of claim 5, wherein the step (d) comprises the steps of: (d-1) extracting the features from each of the concept names stored in the concept name database according to a token, and (d-2) constituting the rule by combining the tokens whose features are extracted, calculating weight value of the constituted rule, filtering the rules with their weight values, and storing the filtered rules in the rule database (page 211-213, *lexemes (single names) and their inflectional variants (keyterms) are determined for a pair of semantic types (concept names). Then*

*discharge summaries are examined to determine which semantic type is used more frequently (weights). The semantic type that is used more frequently is preferred (filtered), and thus assigned to the lexeme and inflectional variants in the semantic lexicon).*

7. As per claim 10, **Johnson** discloses the method of claim 9, wherein in the step (d-1), the feature of the tokens of each of the concept names stored in the concept name database is extracted using the features of the category keyterm, the single name and a capital letter expression, an alphanumeric, a special character, a preposition or conjunction, which are features defined to reflect characteristics of the biological named entity, and a subtype of each of the features (page 211-213, *lexemes (single names) and their inflectional variants (keyterms) are determined for a pair of semantic types (concept names) by matching the semantic lexicon to the metathesaurus. Each lexeme and its variant is matched using first word or letter uppercase, numbers in brackets, a NOS (not otherwise specified) character, and the first preposition in the head noun).*

8. As per claim 11, **Johnson** discloses the method of claim 9, wherein the step (d-2) comprises the steps of: receiving the result in which the concept name is tokenized and the features are extracted at the step (d-1), and creating the rules as many as the number of combinations of subtypes according to the subtypes of the features of the token; and calculating appearance distribution of the rule in each category on all the

created rules, filtering the rules with the threshold, and constructing the rule database (page 211-213, *lexemes (single names) and their inflectional variants (keyterms) are determined for a pair of semantic types (concept names). Then discharge summaries are examined to determine which semantic type is used more frequently (appearance distribution). The semantic type that is used more frequently is preferred (filtering), and thus assigned to the lexeme and inflectional variants in the semantic lexicon.*

9. As per claim 12, **Johnson** discloses the method of claim 5, wherein the steps (f) and (g) comprises the steps of: (f-1) extracting nouns and noun phrases, which are candidate named entities, from the inputted literature; (g-1) extracting features of each token of a candidate named entity; (g-2) combining the features extracted from each of the tokens of the candidate named entity, and creating the rule used to determine the candidate named entity; (g-3) comparing the created rule with the rules stored in the rule database; and (g-4) determining the final semantic category of the candidate named entity (page 211, Corpus Matching, and page 216, Results, *Contiguous word sequences were extracted from a corpus of discharge summaries and matched against the semantic lexicon. The semantic lexicon contains lexemes and semantic types combined to make a rule; therefore a rule must have been extracted from the literature in order to compare it to the semantic lexicon. In addition, preference rules (weight) are applied to the corpus to determine the semantic type.*

10. As per claim 14, **Johnson** discloses the method of claim 12, wherein in the step (g-4), the final semantic category of the candidate named entity is determined using weight values of existing rules extracted at the step (g-3) and a heuristic used to determine a category of the named entity, and outputted as a result of recognizing the named entity (page 211, Corpus Matching, and page 216, Results, *Contiguous word sequences were extracted from a corpus of discharge summaries and matched against the semantic lexicon. In addition, preference rules (weight) are applied to the corpus, which the system used to determine the semantic type.*

11. As per claim 15, **Johnson** discloses the method of claim 1, wherein the candidate named entities are nouns and nouns phrases (page 214, table 11, *examples of lexemes include 'left arm', 'right arm', as well as 'blood', 'aspirin', etc.)*

### ***Claim Rejections - 35 USC § 103***

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

Claim 13 is rejected under 35 U.S.C. 103(a) as being unpatentable over **Johnson** in view of **Veale** (6,584,470).

12. As per claim 13, **Johnson** discloses the method of claim 12, however **Johnson** does not disclose wherein in the step (g-3), existing rules suitable to determine the candidate named entity are extracted an existing rule by comparing the rule used to determine the candidate named entity with the rules stored in the rule database in manners of exact match, partial match and nested match. **Veale** discloses a system for named entity extraction for answering natural language questions (Abstract). In **Veale**, a four-pass search is performed where each pass performs a matching algorithm with different degrees of broadness. The first pass determines an exact match, passes two and three use synonym information to determine exact and partial matches, while pass four determines partial matches (column 20 lines 13-30).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to compare the rule used to determine the candidate named entity with the rules stored in the rule database in manners of exact match, partial match and nested match in **Johnson**, since it would enable the system to utilize different elements of lexical knowledge for each match and allow the use to control a trade off between system accuracy and real-time performance.

#### ***Allowable Subject Matter***

13. Claims 6-8 are objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

***Conclusion***

**THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire **THREE MONTHS** from the mailing date of this action. In the event a first reply is filed within **TWO MONTHS** of the mailing date of this final action and the advisory action is not mailed until after the end of the **THREE-MONTH** shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than **SIX MONTHS** from the mailing date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Dorothy Sarah Siedler whose telephone number is 571-270-1067. The examiner can normally be reached on Mon-Thur 9:30am-5:30pm.

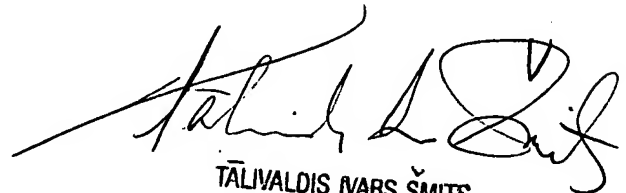
If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on 571-272-7602. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Application/Control Number:  
10/777,072  
Art Unit: 2626

Page 14

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

DSS



TĀLIVALDIS NARS ŠMITS  
PRIMARY EXAMINER